Chapter 10 Hypothesis Testing: Categorical Data

Section 10.1 Introduction (Discrete Variable Example)

- Example 10.1 **Cancer** Suppose we are interested in the association between the use of oral contraceptives (OC use) and the 1-year incidence of cervical cancer from Jan. 1,1988 to Jan. 1, 1989. Women who are disease-free on Jan. 1, 1988, are classified into two OC-use categories as of that date: ever users and never users. We are interested in whether or not the 1-year incidence of cervical cancer is different between ever users and never users. Hence, this is a two-sample problem comparing two binomial proportions, and the t-test methodology in Chapter 8 cannot be used because the outcomes variable, the development of cervical cancer, is a discrete variable with two categories (yes/no), not a continuous variable.

Section 10.1 Introduction (Comparing More Than Two Binomials)

- Example 10.2 **Cancer** Suppose the OC users in Ex. 10.1 are subdivided into "heavy" users, who have used the pill for 5 years or more, and "light" users who have used the pill for less than 5 years. We may be interested in comparing 1-year cervical cancer incidence rates among heavy users, light users, and nonusers. In this problem, three binomial proportions are being compared, and we need to consider methods comparing more than two binomial proportions.

Section 10.1 Introduction (Goodness of Fit Tests)

- Ex. 10.3 **Infectious Disease** The fitting of a probability model based on the Poisson distribution to the random variable defined by the annual number of deaths dues to polio in the United States during the period 1968-1976 has been discussed, as shown in Table 4.8. We want to develop a general procedure for testing the goodness of fit of this and other probability models on actual sample data.

Section 10.2 Two-Sample Test for Binomial Proportions

- **Ex. 10.4 Cancer** A hypothesis has been proposed that breast cancer in women is caused in part by events that occur between the age at menarche (the age when menstruation begins) and the age at first childbirth. In particular, the hypothesis is that the risk of breast cancer increases as the length of this time interval increases. If this theory is correct, then an important risk factor for breast cancer is age at first birth. This theory would explain in part why breast-cancer incidence seems to be higher for women in the upper socioeconomic groups, because they tend to have their children relatively late. An international study was set up to test this hypothesis [1]. Breast-cancer cases were identified among women in selected hospitals in the United States, Greece, Yugoslavia ,Brazil, Taiwan, and Japan. Controls were chosen from women of comparable age who were in the hospital at the same time as the cases but who did *not* have breast cancer. All women were asked about their age at first birth. The set of women with at least one birth was arbitrarily divided into two categories:(1) women whose age at first birth was <= 29, and (2) women whose age at first birth was >= 30. The following results were found among women with at least one birth: 683 out of 3220 (21.2%) women with breast cancer (case women) and 1498 out of 10,245 (14.6%) women without breast cancer (control women) had an age at first birth >= 30. How can we assess whether this difference is significant or simply due to chance?

Section 10.2 Two Sample Test for Binomial Proportions

- Eq. 10.3 **Two-Sample Test for Binomial proportions (Normal Theory Test)** To test the hypothesis $H_0: p_1 = p_2$ versus $H_1: p_1 != p_2$, where the proportions are obtained from two independent samples, use the following procedure:

1. Compute the test statistic

$$z = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$where\ \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \hat{q} = 1 - \hat{p}$$

Section 10.2 Two Sample Test for Binomial Proportions

- Eq 10.3 (Continued)

and x_1, x_2 are the number of events in the first and second samples, respectively.

2. For a two-sided level a test,
   1. If $z > z_{1-\alpha/2}$ then reject H_0
   2. If $z <= z_{1-\alpha/2}$ then accept H_0

3. The approximate p-value is given by $p = 2 * [1 - \Phi(z)]$

4. Use this test only when the normal approximation to the binomial distribution is valid for each of the two samples, that is, when

$$n_1 \hat{p}\hat{q} \geq 5 \; and \; n_2 \hat{p}\hat{q} \geq 5$$

prop.test in R - I

Description:

'prop.test' can be used for testing the null that the proportions (probabilities of success) in several groups are the same, or that they equal certain given values.

Usage:

```
prop.test(x, n, p = NULL,
          alternative = c("two.sided", "less", "greater"),
          conf.level = 0.95, correct = TRUE)
```

prop.test in R - II

Arguments:

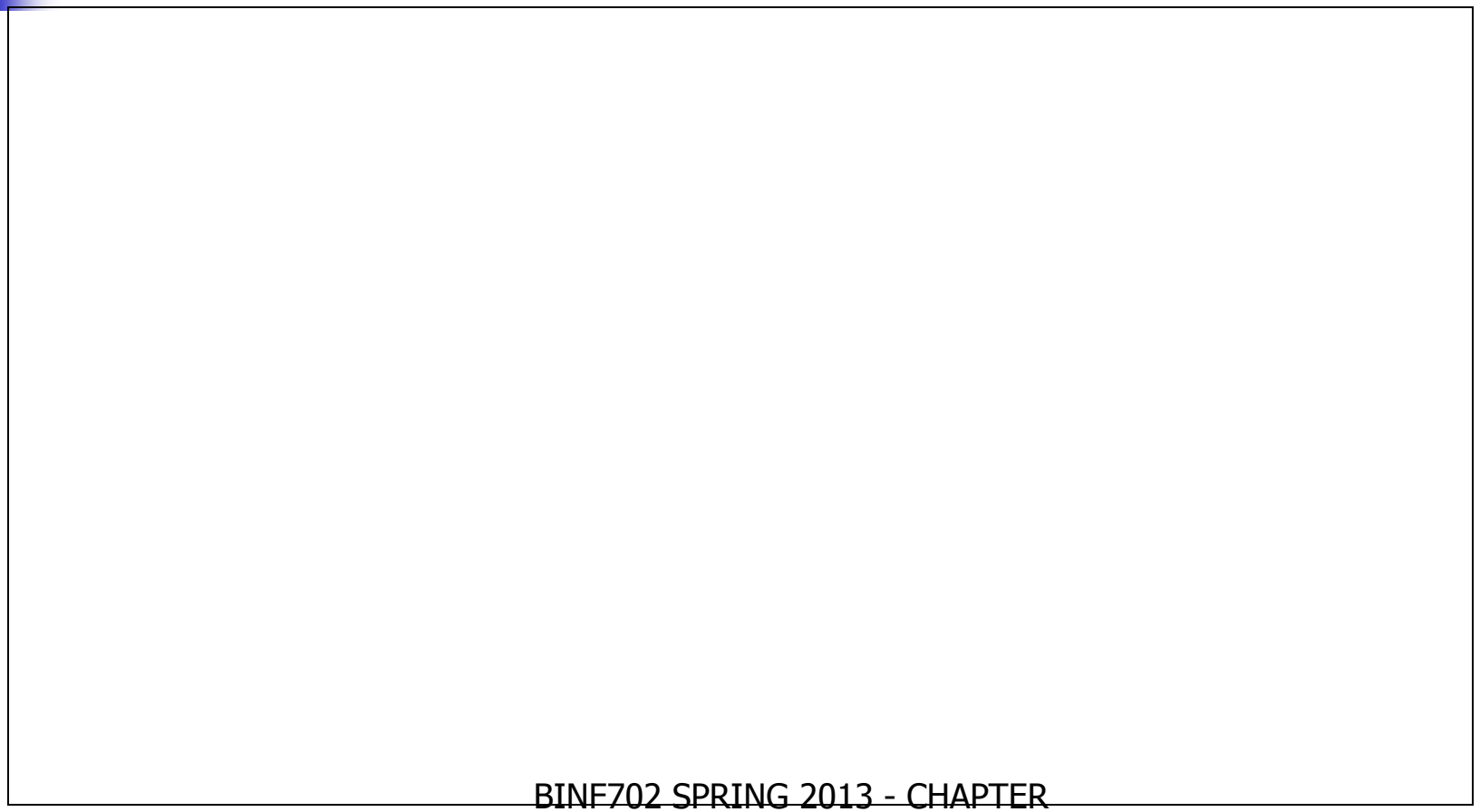
x: a vector of counts of successes or a matrix with 2 columns giving the counts of successes and failures, respectively.

n: a vector of counts of trials; ignored if 'x' is a matrix.

p: a vector of probabilities of success. The length of 'p' must be the same as the number of groups specified by 'x', and its elements must be greater than 0 and less than 1.

correct: a logical indicating whether Yates' continuity correction should be applied.

Example 10.4 in R



Section 10.2 Two Sample Test for Binomial Proportions

- **Pedagogical Example 1** – In a 1985 study of effectiveness of streptokinase in the treatment of patients who have been hospitalized after myocardial infarction, 9 of 199 males receiving streptokinase and 13 of 97 males in the control group died within 12 months. Test for differences in 12-month mortality between the two groups.

Section 10.2 Two Sample Test for Binomial Proportions

- **Pedagogical Example 2** – A case-control study was performed among 2982 cases, 5782 controls, from 10 geographic areas of the United States and Canada. The cases were newly diagnosed cases of bladder cancer in 1977-1978 obtained from cancer registries; the control group was a random sample of the population of the 10 study areas with a similar age, sex, and geographic distribution. The purpose of the study was to investigate the possible association between the incidence of bladder cancer and the consumption of alcoholic beverages. Not all subjects responded. Suppose 574/2388 are drikers and that 980/4660 are not.

Section 10.2.2 Contingency-Table Method

- Ex. 10.7 **Cancer** Suppose all women with at least one birth in the international study in Ex. 10.4 are classified as either cases or controls and with age at first birth as either <= 29 or >= 30. The 4 possible combinations are displayed in Table 10.1

	Age at First Birth		
Status	>= 30	<= 29	Total
Case	683	2537	3220
Control	1498	8747	10245
Total	2181	11284	13465

Section 10.2.2 Contingency-Table Method

- Def 10 .1 – **A 2 x 2 contingency table** is a table composed of two rows cross-classified by two columns. It is an appropriate way to display data that can be classified by two different variables, each of which has only two possible outcomes. One variable is arbitrarily assigned to the rows and the other to the columns. Each of the four cells represents the number of units (women, in the previous example) with a specific value for each of the two variables. The cells are sometimes referred to by number, with the (1,1) cell being in the first row the first column, the (1,2) cell in the first row second column, the (2,1) cell being in the second row and first column, and the (2,2) cell being the cell in the second row second column. The observed number of units in the four cells re likewise referred to as O_{11}, O_{12}, O_{21}, O_{22} respectively. Furthermore it is customary to total
  - The number of units in each row and display them in the right margins, which are called row marginal totals or row margins.
  - The number of units in each column and display them in the bottom margins, which are called column totals or column margins.
  - The total number of units in the four cells, which is displayed in the lower right-hand corner of the table is called the grand total.

Section 10.2.2 Contingency-Table Method

- Sampling designs that lend themselves to the contingency-table design
  - Test for homogeneity of binomial proportions
    - One of the margins is fixed (e. g. rows) and the number of successes in each row is a random variable
  - Testing the independence of two characteristics in the same sample when neither characteristic is particularly appropriate as a denominator. In this setting both sets of margins are assumed to be fixed.

Section 10.2.2 Contingency-Table Method

- Ex. 10.9 – The food-frequency questionnaire is widely used to measure dietary intake. A person specifies the number of servings consumed per week of each of many different food items. The total nutrient composition is then calculated from the specific dietary components of each food item. One was to judge how well a questionnaire measures dietary intake is by its reproducibility. To assess reproducibility the questionnaire is administered at two different times to 50 people and the reported nutrient intakes of the two questionnaires are compared. Suppose dietary cholesterol is quantified on each questionnaire as high if x > 300 mg/day and normal otherwise. The contingency table in Table 10.3 is a natural way to compare the results of the two surveys. Notice that there is some relationship between the two reported measures of dietary cholesterol for the same person. More specifically, we want to assess how unlikely it is that 15 women will report high dietary cholesterol intake on both questionnaires, given that 20 out of 50 women, report high intake on the first questionnaire and 24/50 women report high intake on the second questionnaire. The test is referred to as a test of independence or a test of association between the two characteristics.

	Second Frequency Food Questionnaire		
First Food Frequency Questionnaires	High	Normal	Total
High	15	5	20
Normal	9	21	30
Total	24	26	50

Section 10.2.2 Contingency-Table Method

- Eq 10.4 – **Computation of Expected Values for 2 x 2 Contingency Tables** The expected number of units in the (i, j) cell, which is usually denoted E_{ij}, is the product of the i-th row margin multiplied by the j-th column margin, divided by the grand total.

- Consider Example 10.10.

Section 10.2.2 Contingency-Table Method

- Eq. 10.5 – **Yates-Corrected Chi-Square Test for a 2 x 2 Contingency Table** Suppose we wish to test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 != p_2$ using a contingency-table approach, where O_{ij} represents the observed number of units in the (i, j) cell and E_{ij} represents the expected number of units in the (i, j) cell

$$X^2 = \frac{\left(|O_{11} - E_{11}| - .5\right)^2}{E_{11}} + \frac{\left(|O_{12} - E_{12}| - .5\right)^2}{E_{12}} + \frac{\left(|O_{21} - E_{21}| - .5\right)^2}{E_{21}} + \frac{\left(|O_{22} - E_{22}| - .5\right)^2}{E_{22}}$$

Section 10.2.2 Contingency-Table Method

- Eq 10.5 (Continued)

which under H_0 approximately follows a χ_1^2 distribution.

2. For a level a test, reject H_0
   1. if $X^2 > \chi^2_{1,1-\alpha}$ and
   2. accept H_0 if $X^2 <= \chi^2_{1,1-\alpha}$
3. The approximate p-value is given by the area to the right of X^2 under a χ_1^2 distribution.
4. Use this test only if none of the four expected values is less than 5.

Section 10.2.2 Contingency-Table Method

- The test procedures in equations 10.3 and 10.5 results in the same decision and p-value. prop.test implements 10.5 but we can use it for problems that are formulated as 10.3 problems.

Section 10.2.2 Contingency-Table Method

Ex. 10.14

Section 10.2.2 Contingency-Table Method

- **Pedagogical Example # 1** In a 1985 study of effectiveness of streptokinase in the treatment of patients who have been hospitalized after myocardial infarction, 9 of 199 males receiving streptokinase and 13 of 97 males in the control group died within 12 months. Test for differences in 12-month mortality between the two groups. Use a contingency table approach with prop.test.

Section 10.2.2 Contingency-Table Method

- **Pedagogical Example #2**– A case-control study was performed among 2982 cases, 5782 controls, from 10 geographic areas of the United States and Canada. The cases were newly diagnosed cases of bladder cancer in 1977-1978 obtained from cancer registries; the control group was a random sample of the population of the 10 study areas with a similar age, sex, and geographic distribution. The purpose of the study was to investigate the possible association between the incidence of bladder cancer and the consumption of alcoholic beverages.

Section 10.3 Fisher's Exact Test

- **Ex. 10.16 Cardiovascular Disease** Suppose we wish to investigate the relationship between high salt intake and the occurrence of death from cardiovascular disease (CVD). Groups of high- and low-salt users could be identified and followed over a long period of time to compare the relative frequency of death from CVD in the two groups in contrast, a much less expensive study would involve looking at death records, separating the CVD deaths from the non-CVD deaths and then asking a close relative (such as a spouse) about the dietary habits of the deceased, and comparing salt intake of CVD deaths and non-CVD deaths.

- The normal methods of the previous section are not applicable due to the small sample size.

- This type of study is a retrospective study. It is difficult to perform in many cases but it will almost always be less expensive than the former type of study, a prospective study.

Section 10.3 Fisher's Exact Test

- **Ex. 10.17 Cardiovascular Disease, Nutrition** Suppose a retrospective study is done among men aged 50-54 in a specific county over a 1-month period. The investigators attempt to include approximately an equal number of men who died from CVD (the cases) and men who died from other causes (the controls). It is found that of 35 people who died from CVD, 5 were on a high-salt diet before they died, whereas of 25 people who died from other causes, 2 were on such a diet. These data, presented in Table 10.9, are in the form of a 2 x 2 contingency table, and thus methods of Section 10.2.2 may be applicable. The expected values for this table are two small for the methods of the previous section to be applicable.

	Type Of Diet		
Cause of Death	High salt	Low salt	Total
Non-CVD	2	23	25
CVD	5	30	35
Total	7	53	60

Section 10.3 Fisher's Exact Test

- **General Layout for Fisher exact test example**

	Type Of Diet		
Cause of Death	High salt	Low salt	Total
Non-CVD	a	b	a+b
CVD	c	d	c+d
Total	a+c	b+d	n

- **Eq. 10.7 Exact Probability of Observing a Table with Cells a, b, c, d**

$$P(a,b,c,d) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Section 10.3 Fisher's Exact Test

- Eq. 10.10 **Enumeration of All Possible Tables with the Same Margins as the Observed Table**

1) Rearrange the rows and columns of the observed table so that the smaller row total is in the first row and the smaller column total is in the first column. (WOLOG assume a, b, c, d arrangement)
2) Start with the table with 0 in the (1,1) cell. The other cells in this table are then determined from the row and columns margins. Indeed, to maintain the same row and column margins as the observed table, the (1,2) element must be a a + b, the (2, 1) cell must be a + c, and the (2, 2) element must be (c + d) – (a + c) = d – a.
3) Construct the next table by increasing the (1, 1) cell by 1 (i. e. from 0 to 1), decreasing the (1, 2) and (2, 1) cells by 1, and increasing the (2, 2) cell by 1.
4) Continue increasing and decreasing the cells by 1 , as in step 3, until one of the cells is 0; a which point all possible tables with the given row and column margins have been enumerated. Each table in the sequence of tables is referred to by its (1, 1) element. Thus, the first table is the 0 table, the next table is the 1 table and so one.

Section 10.3 Fisher's Exact Test

- Ex. 10.19

2	23
5	30

0	25
7	28

1	24
6	29

2	23
5	30

3	22
4	31

4	21
3	32

5	20
2	33

6	19
1	34

7	18
0	35

Section 10.3 Fisher's Exact Test

- Equation 10.11 **Fisher's Exact Test: General Procedure and Computation of p-Value** to test the hypothesis H_0: $p_1 = p_2$ versus H_1:p_1 != p_2, where the expected value of at least one cell is < 5 when the data are analyzed in the form of a 2 x 2 contingency table, use the following procedure.

1) Enumerate all possible tables with the same row and column margins as the observed table, as shown in eq. 10.10.
2) Compute the exact probability of each table enumerated in step 1, using either the computer or the methods in eq. 10.7.
3) Suppose that the observed table is the a table and that the last table enumerated is the k table.
    1) To test the hypothesis H_0:$p_1 = p_2$ versus H_1: p_1 != p_2, the p-value = 2 * min([P(0) + P(1) + … + P(a), P(a) + P(a+1) + … + P(k), .5].
    2) To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 < p_2$, the p-value = P(0) + P(1) + … P(a).
    3) To test the hypothesis H_0: $p_1 = p_2$ versus H_1:$p_1 > p_2$, the p-value = P(a) + P(a+1) + … + P(k).

- For each of these three alternative hypotheses, the p-value can be interpreted as the probability of obtaining a table as extreme as or more extreme than the observed table.

Section 10.3 Fisher's Exact Test

- **The Fisher Exact Test in R**
- Description
  - Performs Fisher's exact test for testing the null of independence of rows and columns in a contingency table with fixed marginals.
- Usage
  - fisher.test(x, y = NULL, workspace = 200000, hybrid = FALSE, control = list(), or = 1, alternative = "two.sided", conf.int = TRUE, conf.level = 0.95) Arguments
  - x either a two-dimensional contingency table in matrix form, or a factor object.
  - y a factor object; ignored if x is a matrix.

Section 10.3 Fisher's Exact Test

- Example 10.20 in R (Note that p-value obtained using R matches the value at the bottom of page 407)

Section 10.3 Fisher's Exact Test

- **Pedagogical Example # 1** In the streptokinase study of our previous pedagogical example 2 of 15 females receiving streptokinase and 4 of 19 females in the control group died within 12 months. Use Fisher's test to test for differences in the mortality rates between these two groups.

Section 10.3 Fisher's Exact Test

- Pedagogical Example # 2 – Consider a study of the relationship between early age at menarche and breast-cancer incidence. We select 50 pre-menopausal breast-cancer cases and 50 pre-menopausal age-matched controls. We find that 5 of the cases have an age at menarche < 11 years, and 1 control has an age at menarche < 11 yrs.. Is this a significant finding.

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Ex. 10.21 – Suppose we want to compare different chemotherapy regimens for breast cancer after mastectomy. The two treatment groups should be as comparable as possible on other prognostic factors. To accomplish this goal, a matched study is set up such that a random member of each matched pair gets treatment A (chemotherapy) perioperatively (i.e. 1 week after mastectomy) and for an additional 6 months, whereas the other member gets treatment B (chemotherapy only perioperatively). The patients are followed for 5 years, with survival the outcome variable. The data are displayed in a 2 x 2 table as shown in Table 10.13.
- We can't use previous tests due to the dependency of the data.

	Outcome		
Treatment	Survive for 5 years	Die within 5 years	Total
A	526	95	621
B	515	106	621
Total	1041	201	1242

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Recasting our data in terms of matched pairs
- This table emphasizes the dependence of the two samples
  - P(B member of pair survived|A member of pair survived) = 510/526 while P(B member of pair survived|A member died) = 5/95.
  - Under the assumption of independence these number should agree

	Outcome Of Treatment B Patient		
Outcome of Treatment A Patient	Survive for 5 years	Die within 5 years	Total
Survive for 5 years	510	16	526
Die within 5 years	5	90	95
Total	515	106	621

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Def. 10.2 – A concordant pair is a matched pair in which the outcome is the same for each member of the pairs.

- Def. 10.3 – A discordant pair is a matched pair in which the outcomes are different for the members of the pair.

- Def. 10.4 – A type A discordant pair is a discordant pair in which the treatment of A member of the pair has the event and the treatment of B member of the pair does not. Similarly, a type B discordant pair is a discordant pair in which the treatment B member of the pair has the even and the treatment A member does not.

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Eq. 10.12 – McNemar's Test for Correlated Proportions - Normal

McNemar's Test for Correlated Proportions—Normal-Theory Test

(1) Form a 2 × 2 table of matched pairs, where the outcomes for the treatme A members of the matched pairs are listed along the rows and the outcom for the treatment B members are listed along the columns.

(2) Count the total number of discordant pairs (n_D) and the number of type discordant pairs (n_A).

(3) Compute the test statistic

$$X^2 = \left(\left| n_A - \frac{n_D}{2} \right| - \frac{1}{2} \right)^2 \Big/ \left(\frac{n_D}{4} \right)$$

An equivalent version of the test statistic is also given by

$$X^2 = \left(\left| n_A - n_B \right| - 1 \right)^2 \Big/ \left(n_A + n_B \right)$$

(4) For a two-sided level α test,

if $X^2 > \chi^2_{1,1-\alpha}$

then reject H_0;

if $X^2 \leq \chi^2_{1,1-\alpha}$

then accept H_0.

(5) The exact p-value is given by $p\text{-value} = Pr\left(\chi^2_1 \geq X^2\right)$

(6) Use this test only if $n_D \geq 20$.

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

McNemar's test in R.

Description

Performs McNemar's chi-squared test for symmetry of rows and columns in a two-dimensional contingency table.

Usage

mcnemar.test(x, y = NULL, correct = TRUE) **Arguments**

X either a two-dimensional contingency table in matrix form, or a factor object.

y a factor object; ignored if x is a matrix.

correct a logical indicating whether to apply continuity correction when computing the test statistic.

Details

The null is that the probabilities of being classified into cells [i,j] and [j,i] are the same.

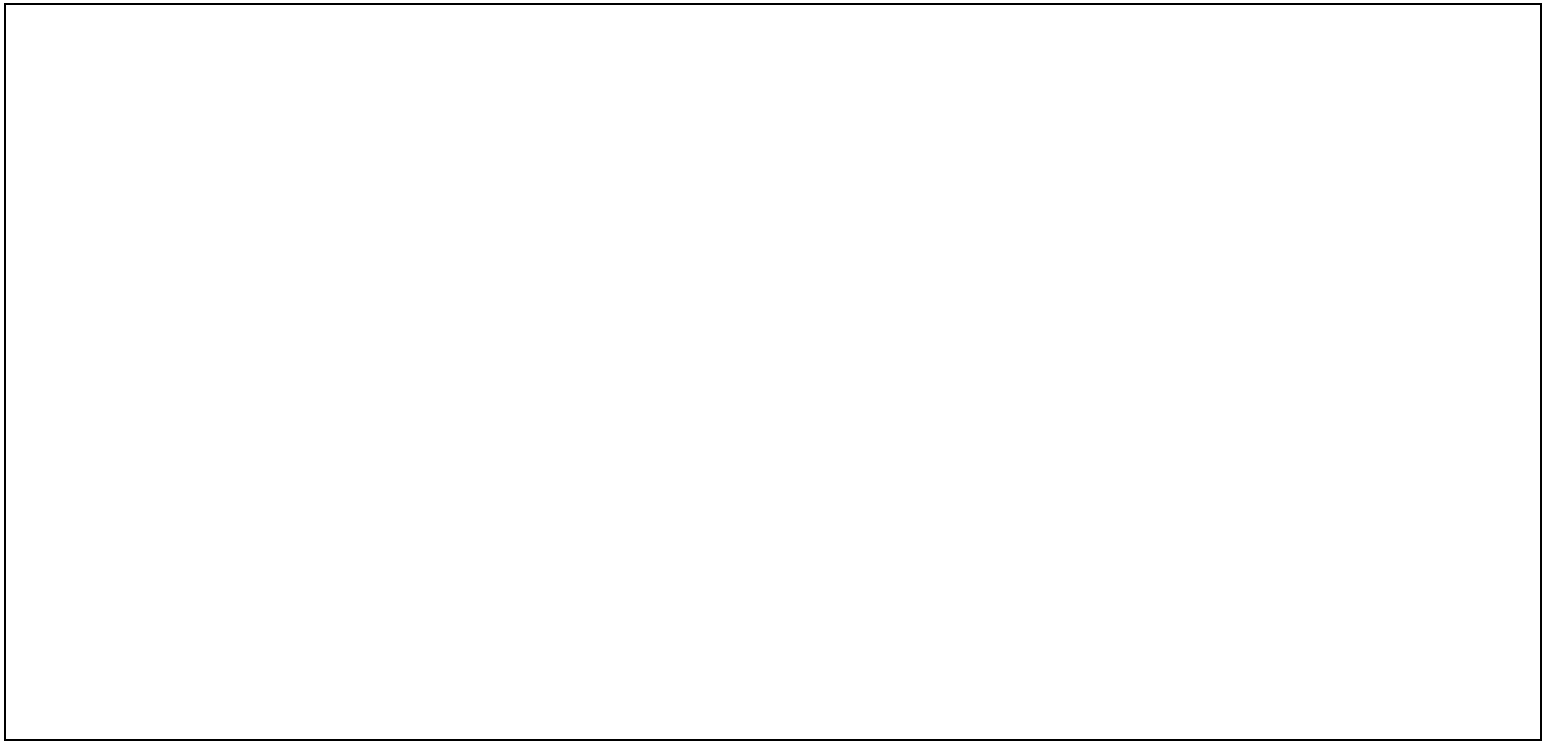
If x is a matrix, it is taken as a two-dimensional contingency table, and hence its entries should be nonnegative integers. Otherwise, both x and y must be vectors of the same length.

Incomplete cases are removed, the vectors are coerced into factor objects, and the contingency table is computed from these.

Continuity correction is only used in the 2-by-2 case if correct is TRUE.

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test) (Ex. 10.24)

- Ex. 10.24



Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Pedagogical Example 1 A case-control study was performed to study the relationship between source of drinking water during prenatal period and congenital malformations. Case mothers are those with malformed infants in a registry in Australia between 1951 and 1979. Controls were individually matched by hospital, maternal age (+- 3 years), and date of birth (+- 1 month). The suspected casual agent was groundwater nitrates. The following 2 x 2 table was obtained relating case-control status to the source of drinking water.

	Source of	Water	
	Ground water	Rain water	
Cases	162	56	218
Controls	123	95	218
Total	285	151	436

Section 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

- Pedagogical Example 1 (Continued)

Case	Control	Frequency
+	+	101
+	-	61
-	+	22
-	-	34

Section 10.4.2 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test) (Exact Case)

- Eq. 10.13McNemar's Test for Correlated Proportions – Exact Test

1) Follow the procedure in step 1 Eq. 10.12
2) Follow the procedure in step 2 in Eq. 10.12
3) $$p = 2 * \sum_{k=0}^{n_A} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D} \quad if \ n_A < n_D / 2$$

$$p = 2 * \sum_{k=n_A}^{n_D} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D} \quad if \ n_A > n_D / 2$$

$$p = 1 \, if \ n_A = n_D / 2$$

4) This test is valid for any number of discordant pairs (n_D) but is particularly useful for $n_D < 20$, when the normal-theory test in Eq. 10.12 cannot be used.

Section 10.4.2 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test) (Exact Case)

- Ex. 10.25 in R (The p-value is slightly different than the one reported in the text).

Section 10.4.2 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test) (Exact Case)

- Pedagogical Example 1 (Revisited) – The data of our previous example was also analyzed separately by season of birth. The following exposure data are presented in a 2 x 2 table of case exposure status by control exposure status for spring births.

	Control		
		+	-
Case	+	30	14
	-	2	10

Section 10.5 Estimation of Sample Size and Power Comparing Two Binomial Proportions

$$n_1 = \frac{\left[\sqrt{\bar{p}\bar{q}\left(1+\frac{1}{k}\right)}z_{1-a/2} + \sqrt{p_1q_1 + \frac{p_2q_2}{k}}\, z_{1-\beta}\right]^2}{\Delta^2}$$

$$n_2 = kn_1$$

where $p_1, p_2 =$ projected true probabilities of success in the two groups

$$q_1, q_2 = 1-p_1, 1-p_2$$

$$\Delta = |p_2 - p_1|$$

$$\bar{p} = \frac{p_1 + kp_2}{1+k}$$

$$\bar{q} = 1-\bar{p}$$

- Eq. 10.14 **Sample Size Needed to Compare Two Binomial Proportions Using a Two-Sided Test with Significance level α and Power 1 – β, Where One Sample (n_2) is k times as Large as the other Sample (n_1) (Independent Sample Case)** To test the hypothesis H_0: $p_1 = p_2$ versus H_1:p_1 != p_2 for the specific alternative $|p_1 - p_2| = \Delta$, with a significance level α and a power 1 – β the following sample size is required

Section 10.5 Estimation of Sample Size and Power Comparing Two Binomial Proportions

- Eq. 10.15 Power Achieved in Comparing Two Binomial Proportions Using a Two-Sided Test with Significance Level α and Samples of Size n_1 and n_2 (Independent Sample Case) To test the hypothesis H_0: $p_1 = p_2$ versus H_1:p_1 != p_2 for the specific alternative $|p_1 - p_2| = \Delta$, with a significance level α compute

$$Power = \Phi\left[\frac{\Delta}{\sqrt{p_1 q_1 / n_1 + p_2 q_2 / n_2}} - z_{1-\alpha/2}\frac{\sqrt{\bar{p}\bar{q}\left(1/n_1 + 1/n_2\right)}}{\sqrt{p_1 q_1 / n_1 + p_2 q_2 / n_2}}\right]$$

where

$p_1 p_2 =$ projected true probabilities of success in groups 1 and 2 respectively

$q_1, q_2 = 1 - p_1, 1 - p_2$

$\Delta = |p_2 - p_1|$

$$\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$\bar{q} = 1 - \bar{p}$

Section 10.5.2 Estimation of Sample Size and Power Comparing Two Binomial Proportions (Paired Samples)

- Eq. 10.16 **Sample Size Needed to Compare Two binomial Proportions Using a Two Sided Test with Significance Level 1 and Power 1- β (Paired-Sample Case)** If McNemar's test for correlated proportions is used to test the hypothesis H_0: p = ½ versus H_1:p != ½ for the specific alternative p = p_A, where p = the probability that a discordant pair is of type A, then use

$$n = \frac{\left(z_{1-\alpha/2} + 2z_{1-\beta}\sqrt{p_A q_A}\right)^2}{4\left(p_A - .5\right)^2 p_D} \text{ matched pairs}$$

$$\text{or } 2n = \frac{\left(z_{1-\alpha/2} + 2z_{1-\beta}\sqrt{p_A q_A}\right)^2}{2\left(p_A - .5\right)^2 p_D} \text{ individuals}$$

where p_D = projected proportion of discordant pairs all pairs

and p_A = projected proportion of discordant pairs of type A among discordant pairs

Section 10.5.2 Estimation of Sample Size and Power Comparing Two Binomial Proportions (Paired Samples)

- Eq. 10.17 **Power Achieved in Comparing Two Binomial Proportions Using a Two-Sided Tests with Significance Level α (Paired-Sample Case)** If McNemar's test for correlated proportions is used to test the hypothesis H_0: p = ½ versus H_1: p != ½, for the specific alternative p_A, where p = the probability that a discordant pair is of type A,

$$Power = \Phi\left[\frac{1}{2\sqrt{p_A q_A}}\left(z_{\alpha/2} + 2\left|p_A - .5\right|\sqrt{np_D}\right)\right]$$

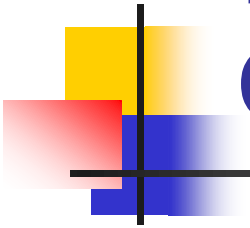
where

n = number of matched pairs

p_D = projected proportion of discordant pairs among all pairs

p_A = projected proportion of discordant pairs of type A among discordant pairs

10.5.3 Sample Size and Power in a Clinical Trial Setting



- This section is left to the reader.

Section 10.6 R x C Contingency Tables

- Def. 10.7 An **R x C contingency table** is a table with R rows and C columns. It displays the relationship between two variables, where the variables depicted in the rows has R categories and the variables depicted in the columns has C categories.
- Ex. 10.33 **Cancer** Suppose we wish to study further the relationship between age at first birth and the development of breast cancer, as given in Example 10.4. In particular, we would like to know if the effect of age at first birth follows a consistent trend, that is (1) more protection for women whose age at first birth is < 20 than for women whose age at first birth is 25-29 and (2) higher risk for women whose age at first birth is >= 35 than for women whose age at first birth is 30-34. The data are presented in Table 10.18, where case-control status, is indicated along the rows and age at first birth has five categories. We wish to test for a relationship between birth and case-control status. How should this be done?

Section 10.6 R x C Contingency Tables

- The data of Example 10.33

	Age	At	First	Birth		
Case-Control Status	<20	20-24	25-29	30-34	>=35	Total
Case	320	1206	1011	463	220	3220
Control	1422	4432	2893	1092	406	10245
Total	1742	5638	3904	1555	626	13465

Section 10.6 R x C Contingency Tables

- Eq. 10.22 **Computation of the Expected Table for an R x C Contingency Table** The expected number of units in the (i, j) cell = E_{ij} = the product of the number of units in the ith row multiplied by the number of units in the jth column, divided by the total number of units in the table.

Section 10.6 R x C Contingency Tables

- Eq. 10.23 **Chi-Square Test for an R x C Contingency Table** To test for the relationship between two discrete variables, where one variable has R categories and the other has C categories, use the following procedure:

1) Analyze the data in the form of an R x C contingency table, where O_{ij} represents the observed number of units in the (i, j) cell.
2) Compute the expected table as shown in eq. 10.22, where E_{ij} represents the expected number of units in the (i,j) cell.
3) Compute the test statistic

$$X^2 = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \ldots + \frac{(O_{RC} - E_{RC})^2}{E_{RC}}$$

Which under H_0 approximately follows a chi-square distribution with (R-1) x (C-1) df

Section 10.6 R x C Contingency Tables

- **Section 10.6 R x C Contingency Tables (cont.)**

4) For a level α test, if
   1) $X^2 > \chi^2_{(R-1)x(C-1),1-\alpha}$ then reject H_0
   2) $X^2 <= \chi^2_{(R-1)x(C-1),1-\alpha}$ then accept H_0
5) The approximate p-value is given by the area to the right of X^2 under a $\chi^2_{(r-1)x(C-1),1-\alpha}$ distribution
6) Use this test only if the following two conditions are satisfied.
   1) No more than 1/5 of the cells should have expected values < 5
   2) No cell should have expected value < 1.

Section 10.6 R x C Contingency Tables

R x C Contingency Tables in R

Description

chisq.test performs chi-squared tests on contingency tables.

Usage

```
chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x), length(x)), simulate.p.value = FALSE,
    B = 2000)
```

Arguments

x a vector or matrix.ya vector; ignored if x is a matrix.

correct a logical indicating whether to apply continuity correction when computing the test statistic.

p a vector of probabilities of the same length of x.

simulate.p.value a logical indicating whether to compute p-values by Monte Carlo simulation.

B an integer specifying the number of replicates used in the Monte Carlo simulation.

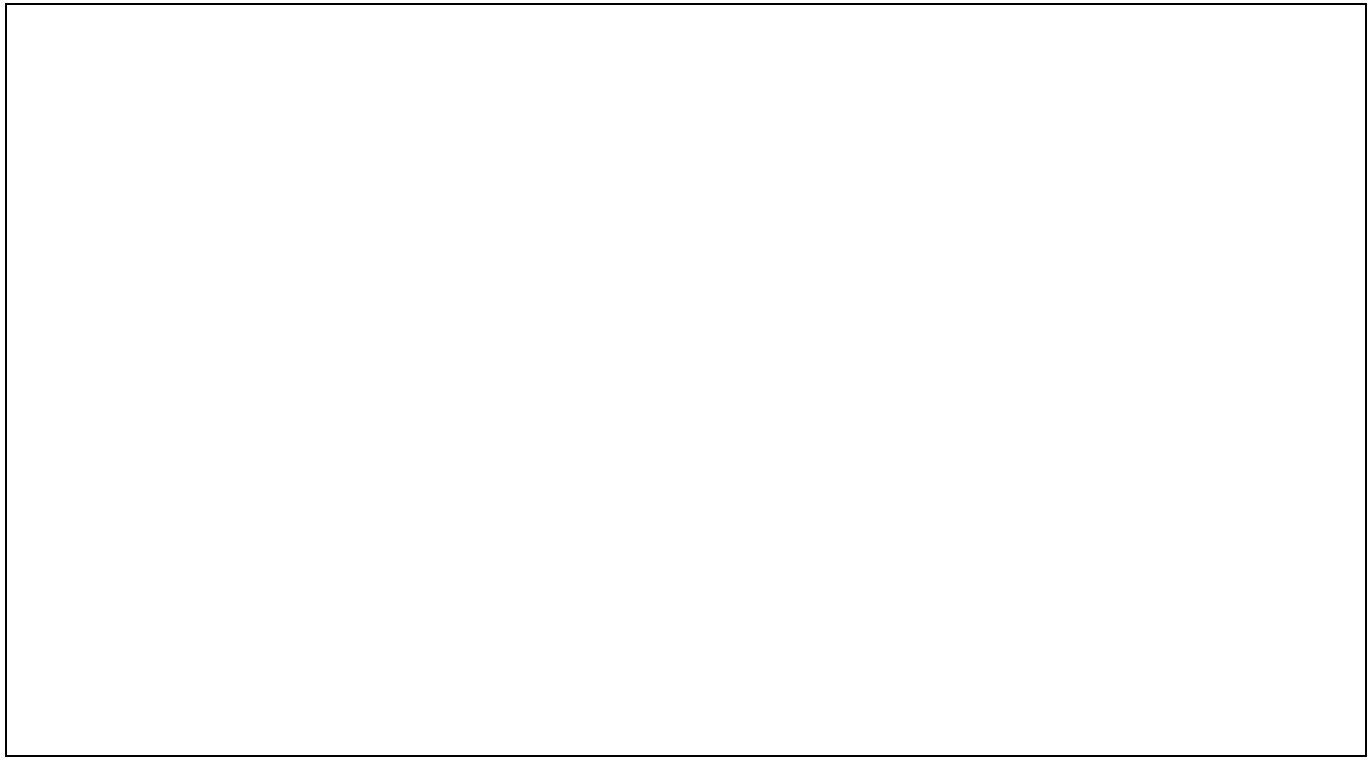
Details

If x is a matrix with one row or column, or if x is a vector and y is not given, x is treated as a one-dimensional contingency table. In this case, the hypothesis tested is whether the population probabilities equal those in p, or are all equal if p is not given.

If x is a matrix with at least two rows and columns, it is taken as a two-dimensional contingency table, and hence its entries should be nonnegative integers. Otherwise, x and y must be vectors or factors of the same length; incomplete cases are removed, the objects are coerced into factor objects, and the contingency table is computed from these. Then, Pearson's chi-squared test of the null that the joint distribution of the cell counts in a 2-dimensional contingency table is the product of the row and column marginals is performed. If simulate.p.value is FALSE, the p-value is computed from the asymptotic chi-squared distribution of the test statistic; continuity correction is only used in the 2-by-2 case if correct is TRUE. Otherwise, if simulate.p.value is TRUE, the p-value is computed by Monte Carlo simulation with B replicates. This is done by random sampling from the set of all contingency tables with given marginals, and works only if the marginals are positive. (A C translation of the algorithm of Patefield (1981) is used.)

Section 10.6 R x C Contingency Tables (Example 10.35)

- Example 10.35 in R



Section 10.6 R x C Contingency Tables

- **Pedagogical Example # 1**
  Patients with heart failure, diabetes, cancer, and lung disease who have various infections from gram-negative organisms often receive aminoglycosides. One of the side effects of aminoglycosides is nephrotoxcity (possible damge to the kidney). A study was performed comparing the nephrotoxity (rise in serum creatine of at least 0.5 mg/dL) for 3 aminoglycosides. The following results were obtained:

	# patients with rise in serum creatinine	Total
Gentamicin	44	121
Tobramycin	21	92
Amikadn	4	16

Section 10.6 R x C Contingency Tables

- **Pedagogical Example # 1 (continued)**

	Antibiotic			
		G	T	A
Nephrotoxicity	+	44	21	4
	-	77	71	12

Section 10.6 R x C Contingency Tables

- **Pedagogical Example # 1 (cont.)**

Section 10.6.2 Chi-Square Test for Trend in Binomial Proportions

- In this section a test is developed to test for a trend in a proportional variable. As part of this procedure a score variable S_i is introduced to correspond to the ith group. The score variable can represent some particular numeric attribute of the group. In other instances, for simplicity, 1 is assigned to the first group, 2 to the second group, …, k to the kth group.

- In this test
  - H_0: p_i are all equal vs. H_1: $p_i = \alpha + \beta S_i$

- Assigning $S_i = i$ our alternative hypothesis takes the form of
  - $P_i = \alpha + \beta_i$

- The implementation of the testing procedure is outlined in Eq. 10.24 pg. 431 of your text.

Section 10.6.2 Chi-Square Test for Trend in Binomial Proportions

```
prop.trend.test

Description:

     Performs chi-squared test for trend in proportions, i.e., a test
     asymptotically optimal for local alternatives where the log odds
     vary in proportion with 'score'.  By default, 'score' is chosen as
     the group numbers.

Usage:

     prop.trend.test(x, n, score = 1:length(x))

Arguments:

       x: Number of events

       n: Number of trials

   score: Group score
```

Section 10.6.2 Chi-Square Test for Trend in Binomial Proportions (Example 10.37)

Example 10.37 in R

Section 10.6.2 Chi-Square Test for Trend in Binomial Proportions

- Pedagogical Example # 1-A study reported that out of 311 people who had attempted to quit smoking, 16 out of 33 with less than a high school education were successful quitters; 47 out of 76 who had finished high school but had not gone to college were successful quitters; 69 out of 125 who attended college but did not finish 4 years of college were successful quitters; and 52 out of 77 who had completed college were successful quitters. Do these data show an association between the number of years of education and the rate of successful quitting?

Section 10.7- Chi-Squared Goodness-of-Fit Test

- Example 10.39 – Diastolic blood pressure measurements were collected at home in a community-wide screening program of 14,736 adults ages 30-69. in East Boston., Massachusetts, as part of a nationwide study to detect and treat hypertensive people. The people in the study were each screened in the home, with two measurements taken at one visit. A frequency distribution of the mean diastolic blood pressure is given in the Table below in 10 mm Hg intervals. How can we assess whether these measurements come from an underlying normal distribution?

Section 10.7- Chi-Squared Goodness-of-Fit Test

Group (mm HG)	Observed Frequency	Expected Frequency	Group	Observed Frequency	Expected Frequency
<50	57	77.9	>=80, <90	4604	4478.5
>=50, <60	330	547.1	>=90, <100	2119	243.1
>=60,<70	2132	2126.7	>=100, <110	659	684.1
>=70,<80	4584	4283.3	>=110	251	107.2
		Total		14,736	14,736

Section 10.7- Chi-Squared Goodness-of-Fit Test

- Eq. 10.26 (Chi-Square Goodness-of-Fit Test) To test for the goodness of fit of a probability model, use the following procedure

1. Divide the data into groups. The considerations for grouping data are similar to those in Section 2.7. In particular, the groups must not be too small, so step 7 is not violated.
2. Estimate the k parameters and the probability models from the data using the methods of Chapter 6.
3. Use the estimates in step 2 to compute the probability phat of obtaining a value within a particular group and the corresponding expected frequency with that group (n*phat), where n is the total number of data points.
4. If O_i and E_i are respectively the observed and the expected number of units within the ith group, then compute

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \cdots + \frac{(O_g - E_g)^2}{E_g}$$

where g = number of groups

Section 10.7- Chi-Squared Goodness-of-Fit Test

5. For a test with a significance level a, if

$$X^2 > \chi^2_{g-k-1,1-\alpha}$$

then reject H_0; if

$$X^2 \leq \chi^2_{g-k-1,1-\alpha}$$

then accept H_0

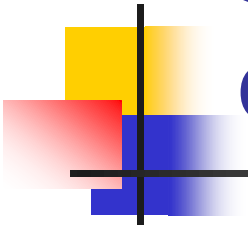
6. The approximate p-value for this test is given by

$$\Pr\left(\chi^2_{g-k-1} > X^2\right)$$

Section 10.7- Chi-Squared Goodness-of-Fit Test

7. Use this test only if
   1. No more than 1/5 of the expected values are < 5.
   2. No expected value is < 1.
8. Note; If the parameters of the probability model were specified a priori without using the present sample data, then k = 0 and $X^2 \sim \chi_{g-1}^2$. We call such a model a externally specified model, as opposed to the internally specified model described in steps 1 through 7.

Section 10.7- Chi-Squared Goodness-of-Fit Test in R



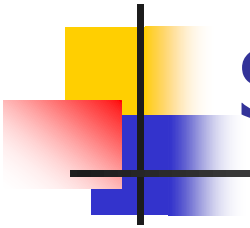
A recipe to hack a chi-square goodness-of-fit test in R.

(1) Cut the data into bins (you can use hist() to do this)

(2) Calculate the expected numbers in each bin using the differences of the CDF (pnorm, pexp, etc.)

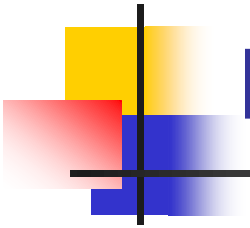
(3) calculate sum((exp-obs)^2/exp)

(4) find the tail probability of the chi-square distribution (pchisq).



Section 10.8 – The Kappa Statistic

- Please read this section on your own.



Homework Chapter 10

- 10.1, 10.2, 10.6, 10.7, 10.8, 10.9, 10.18, 10.20, 10.24, 10.36